The Pennsylvania State University

Department of Statistics

University Park, Pennsylvania

TECHNICAL REPORTS AND PREPRINTS

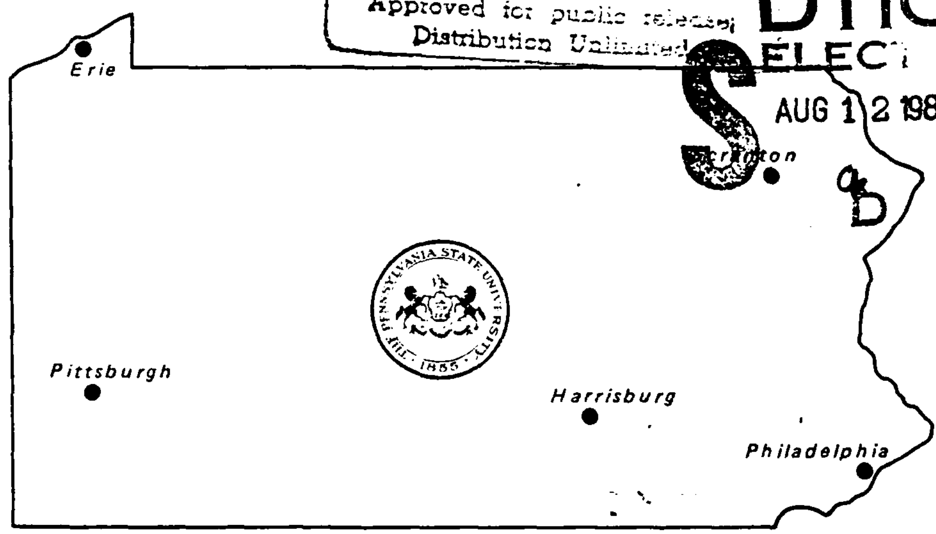
Number 74: August 1988

TWO-SAMPLE INFERENCE FOR MEDIAN SURVIVAL TIMES BASED ON ONE-SAMPLE PROCEDURES FOR CENSORED SURVIVAL DATA

Jane-Ling Wang*
The Wharton School
University of Pennsylvania

and

Thomas P. Hettmansperger**
The Pennsylvania State University



DEPARTMENT OF STATISTICS

The Pennsylvania State University
University Park, PA 16802 U.S.A.

4

TECHNICAL REPORTS AND PREPRINTS

Number 74: August 1988

TWO-SAMPLE INFERENCE FOR MEDIAN SURVIVAL TIMES BASED ON ONE-SAMPLE PROCEDURES FOR CENSORED SURVIVAL DATA

Jane-Ling Wang*
The Wharton School
University of Pennsylvania

and

Thomas P. Hettmansperger**
The Pennsylvania State University

* The work of this author was partially supported by Air Force Grant AFOSR-85-0268.

** The work of this author was partially supported by ONR Contract N00014-80-C0741.

Abstract

Confidence intervals for median survival times are derived for censored survival data. The intervals are obtained by using the quantiles of the Kaplan-Meier product limit estimator and have the same Pitman efficiency as the intervals found by inverting the sign test. Two-sample tests and confidence intervals for the difference in median survival times are then developed based on the comparison of the one-sample confidence intervals. Several methods for choosing the confidence coefficients of the corresponding one-sample confidence intervals are developed. The Pitman efficiencies of these two-sample tests are the same as that of the median test proposed by Brookmeyer and Crowley (1982b). The procedures can also be used for the Behrens-Fisher problem.

Nonparametric two-sample inference procedures are useful in comparing the responses of treatment and control groups. In medical follow up studies the data are usually subjected to censoring. While many of the two-sample procedures have been extended to accommodate censored data we show, in this paper, how to construct two-sample tests and confidence intervals based on one-sample confidence intervals. This method was first discussed in Hettmansperger (1984a) for uncensored data where a confidence interval for the difference in population medians is constructed by subtracting the endpoints of one (one-sample) interval from the opposite endpoints of the other. The test then rejects the null hypothesis of equal medians if the one-sample intervals are disjoint. Hettmansperger used the sign-interval as the one-sample interval which is obtained by inverting the sign-test. In the presence of censoring we modify the sign-interval to the so called quantile-interval whose endpoints are the quantiles of the product-limit estimator of Kaplan and Meier (1958). After deriving the asymptotic properties of the quantile-interval we show, for specified overall level α, three ways to select the

confidence coefficients for the one-sample quantile intervals. For $\alpha = 0.05$ and equal confidence coefficients for both samples, the one-sample confidence coefficients are in the neighborhood of .85 under the proportional hazard model. The procedures are then applied to data from a colorectal cancer clinical trial to compare four treatments.

Key Words: Product-limit estimator, median test, sign test, confidence interval, quantile, asymptotic relative efficiency.

Section 1. Introduction

Nonparametric two-sample tests based on ranks, e.g. median and Wilcoxon tests, have been extended by Gehan (1965), Efron (1967), and Brookmeyer and Crowley (1982b) among others to accommodate censored survival data. While there is a vast literature on general linear rank test procedures with censored observations, two-sample confidence intervals have received much less attention. Nonparametric confidence intervals for the ratio of two scale parameters were derived for randomly right censored data by Wei and Gail (1983) based on the idea of Hodges and Lehmann (1963). Their method can also be applied to obtain confidence intervals for the difference of two location parameters. The confidence sets were computed by a numerical method based on a grid search and may not yield an interval. In this paper we shall focus on the problem of testing and confidence intervals for the difference of two population medians with randomly right censored data. Inferences for other two-sample location and scale parameters will be discussed briefly in section 6.

Following the idea of Hettmansperger (1984a), we shall show how to construct a two-sample test and confidence interval based on one-sample confidence intervals for the median of each population. Let $[L_x, U_x]$ and $[L_y, U_y]$ denote one-sample confidence intervals for the population medians θ_x and θ_y with confidence coefficients γ_x and γ_y, respectively. A confidence interval for $\Delta = \theta_y - \theta_x$, the difference of the population medians, can be obtained as $[L_y - U_x, U_y - L_x]$ with coefficient $1 - \alpha$. The overall coefficient $1 - \alpha$ obviously depends on the respective one-sample coefficients γ_x and γ_y. Bonferoni's inequality can be applied to obtain a bound on the desired overall level but it is usually too conservative. Using one-sample sign-intervals and their asymptotic distributions, Hettmansperger (1984a) showed how to select γ_x and γ_y for a specified α. For example; if one chooses equal coefficients then $\gamma_x = \gamma_y$ is about .84 for $\alpha = .05$. Further, the size α two-sample test of H_0: $\Delta = 0$ vs. H_A: $\Delta \neq 0$ rejects H_0 when the one-sample intervals are disjoint.

In the presence of censoring, the sign-intervals need to be modified. To motivate the construction of a sign-interval with censored data let us first consider the situation when there is no censoring. Let $X_{(1)} \leq \ldots \leq X_{(m)}$ be an ordered sample from a continuous distribution with unique median θ_x. The sign-interval is $[L_x, U_x] = [X_{(d)}, X_{(m-d+1)}]$ with confidence coefficient $\gamma_x = 1-2\,p(S < d)$, where S has a binomial distribution with parameters m and .5 and d is called the depth of the sign-interval. Given any distribution function F(t), define its left continuous inverse to be

$$F^{-1}(p) = \inf\{t: \; F(t) \geq p\} \text{ for } 0 \leq p \leq 1 \tag{1.1}$$

One then notices that $X_{(d)} = \hat{F}_m^{-1}(d/m)$ and $X_{(m-d+1)} = \hat{F}_m^{-1}((m-d+1)/m)$, where $\hat{F}_m(\cdot)$ is the empirical distribution function. In the presence of censoring, the nonparametric counterpart of $\hat{F}_m(\cdot)$ is the product-limit estimator (PLE) $F_m(\cdot)$ proposed by Kaplan and Meier (1958). It is now clear that, for a specified depth d, one can use

$$[L_x, U_x] = [F_m^{-1}(d/m), \; F_m^{-1}((m-d+1)/m)] \tag{1.2}$$

as a confidence interval for the median. The confidence coefficient γ_x will no longer be available from a binomial table and will have to be approximated from a normal table due to the censoring scheme. It should be mentioned here that the confidence interval (1.2) is different from the confidence region R_α based on the sign test in Brookmeyer and Crowley (1982a). For this reason and since (1.2) is obtained from estimates of the quantiles of the true distribution function, hereafter we shall refer to (1.2) as the quantile-interval with depth d. The confidence region R_α need not be an interval so they suggested the use of only the interval part of the region and denoted it by I_α. This truncation of the confidence region to an interval complicates the issue of establishing two-sample inference procedures based on one-sample procedures. We therefore consider the quantile-intervals, and we examine the effect of censoring by comparing our results to the uncensored case studied by Hettmansperger (1984a). One-sample confidence intervals for the median which are based on the bootstrap method were derived by Reid (1981).

Let $[L_y, U_y]$ be the quantile-interval constructed as in (1.2) for an independent y sample with confidence coefficient γ_y. The selection of γ_x and γ_y for a specified α is given in section 3 after deriving the asymptotic distributions of the relevant statistics. Three different ways to specify γ_x and γ_y are given. Under the proportional hazards model some values of γ_x and γ_y are tabulated in section 4. In particular, for $\alpha = 0.05$ and equal coefficients ($\gamma_x = \gamma_y$), the coefficients γ_x and γ_y are in the range of .83 to .88 even for heavy censoring and for many pairs of sample sizes. Our asymptotic results are facilitated by i.i.d. representations of the product-limit estimator and its quantile process due to Lo and Singh (1986). Their results are summarized in section 2. In section 5 we apply the procedures to data from a phase III colorectal cancer clinical trial.

It was shown in Hettmansperger (1984a) that the tests based on sign-intervals have the same efficiency as Moods' (1950) median test. In the presence of censoring, Moods' median test has been extended in several ways by Prentice (1978) and Brookmeyer and Crowley (1982). Gastwirth and Wang (1987) extend the control percentile test which is also a median type test when the specified percentile is the median. All these tests turn out to have the same Pitman efficiency as our test.

Section 2. Preliminaries

Let us first introduce the two-sample location problem with randomly right censored data. Let $X_1^o, X_2^o, \ldots, X_m^o$ and $Y_1^o, Y_2^o, \ldots, Y_n^o$ be two independent random samples from continuous life distribution functions $F^o(t) = L(t - \theta_x)$ and $G^o(t) = L(t - \theta_y)$, respectively, where $L(\cdot)$ is a continuous distribution function with unique median zero. Hence, $G^o(t) = F^o(t - \Delta)$, where $\Delta = \theta_y - \theta_x$. Let $C_1, C_2, \ldots, C_m$ and $D_1, D_2, \ldots, D_n$ denote two independent random samples from arbitrary censoring distribution functions $H(\cdot)$ and $K(\cdot)$, respectively. In the random right censoring model, one observes $\{(X_i, \varepsilon_i), i = 1, \ldots, m\}$ and $\{(Y_j, \delta_j), j = 1, \ldots, n\}$, where $X_i = \min(X_i^o, C_i) = X_i^o \wedge C_i$, $\varepsilon_i = I(X_i = X_i^o)$, $Y_j = \min(Y_j^o, D_j) = Y_j^o \wedge D_j$, $\delta_j = I(Y_j = Y_j^o)$ and $I(\cdot)$ is the indicator function. The censoring times will be assumed to be independent of the corresponding lifetimes, and hence the survival functions of X_i and Y_j are $\bar{F}(t) = \bar{F}^o(t)\,\bar{H}(t)$ and $\bar{G}(t) = \bar{G}^o(t)\,\bar{K}(t)$ respectively, where for any distribution function $W(\cdot)$, its survival function is $\bar{W}(t) = 1 - W(t)$.

Denote the Kaplan-Meier (1958) product-limit estimators (PLE) of F^o and G^o by F_m^o and G_n^o. With proper choice of d_x and d_y to be discussed later, the quantile-intervals for θ_x and θ_y with confidence levels γ_x and γ_y respectively are,

$$[L_x, U_x] = [F_m^{o\,-1}(d_x/m),\ F_m^{o\,-1}((m - d_x + 1)/m)] \tag{2.1}$$

and

$$[L_y, U_y] = [G_n^{o\,-1}(d_y/n),\ G_n^{o\,-1}((n - d_y + 1)/n)]\ , \tag{2.2}$$

where the inverse function $F_m^{o\,-1}$ is defined in (1.1). For specified α, called the overall error rate, we shall show in the next section how to pick the two quantile-intervals so that:

(a) The level α test for $H_0: \Delta = 0$ versus $H_A: \Delta \neq 0$ rejects H_0 whenever the quantile-intervals are disjoint.

(b) The $(1-\alpha)$ confidence interval for Δ is

$$[L_y - U_x, U_y - L_x]. \tag{2.3}$$

The choices of γ_x and γ_y (or equivalently of d_x and d_y) which are given in the next section, are based on the normal approximations of the endpoints of the quantile-intervals. We shall now introduce the background of such approximations which are based on i.i.d. representations of the PLE F_m^o and its quantile function $F_m^{o\,-1}(\cdot)$ due to Lo and Singh (1986).

Let $F_1(t) = P(X_1 \leq t, \varepsilon_1 = 1)$ and $G_1(t) = P(Y_1 \leq t, \delta_1 = 1)$ be the subdistribution functions for the uncensored observations. For positive reals x, t and for ε taking values zero or one, let

$$\xi(x,\varepsilon,t) = \bar{F}^o(t) \left\{ \frac{I(x \leq t, \varepsilon = 1)}{\bar{F}(x)} - \int_0^{x \wedge t} \frac{d\, F_1(s)}{[\bar{F}(s)]^2} \right\} \tag{2.4}$$

Similarly, one can define $\eta(y, \delta, t)$ using instead G^o, G and G_1 in (2.4). The following proposition is an extension of Theorem 1 of Lo and Singh (1986). A proof can be found in Lo, Mack and Wang (1986).

Proposition 1. For any T such that $F(T) < 1$, we have

$$F_m^o(t) - F^o(t) = \frac{1}{m} \sum_1^m \xi(X_i, \varepsilon_i, t) + R_m(t)$$

$$= \bar{\xi}_m(t) + R_m(t) ,$$

where $\sup_{0 \le t \le T} |R_m(t)| = O(\ell n\, m/m)$ a.s. and ℓn denotes the natural logarithm.

Remarks. 1. The $\xi(X_i, \varepsilon_i, t)$'s are i.i.d. with mean 0 and variance

$$[\bar{F}^o(t)]^2 \int_o^t \frac{dF_1(s)}{[\bar{F}(s)]^2} .$$

2. The process $m^{1/2}\, \bar{\xi}_m(t)$ converges weakly to a mean zero Gaussian process $Z(t)$ which has continuous paths with probability one on $D[0,T]$.

The next proposition is a slight variant of the quantile representation in Theorem 2 of Lo and Singh (1986) which is an extension, in the presence of censoring, of Bahadur's representation for the sample quantiles (Serfling 1980, p. 91). The current variant is similar to Ghosh's extension of Bahadur's representation (Serfling 1980, p. 92) as we are concerned with $n^{1/2}$ - asymptotics, and the proof can be obtained along the same lines.

Proposition 2. For any $0 < p < 1$, if the derivative $f^o(t)$ of $F^o(t)$ is continuous and positive at $t = F^{o\ -1}(p)$, we have

$$F_m^{o\ -1}(p) - F^{o\ -1}(p) = -\frac{1}{m} \sum_{i=1}^m \frac{\xi(X_i, \varepsilon_i, F^{o\ -1}(p))}{f^o(F^{o\ -1}(p))} + o_p(m^{-1/2})$$

$$= -\bar{\xi}_m(F^{o\ -1}(p))/f^o(F^{o\ -1}(p)) + o_p(m^{-1/2}) .$$

Similar results can be obtained for the y-sample and we will denote the mean process by $\bar{\eta}_n(t)$.

Section 3. Asymptotic Behavior of Two-sample Tests and Confidence Intervals.

Consider the two-sample location model in Section 2. In this section we shall assume that the distribution L is continuously differentiable near zero with density $\ell(0) > 0$. The asymptotic results in section 2 of Hettmansperger (1984a) will be extended to the censored data case.

Theorem 1. Let Z_x be a positive constant and

$$d_x = m/2 + 0.5 - Z_x m^{1/2}/2 \ . \tag{3.1}$$

Then

$$\begin{aligned} L_x &= F_m^{o\ -1}(d_x/m) = \theta_x - Z_x/[2m^{1/2}\ \ell(0)] - \bar{\xi}_m(\theta_x)/\ell(0) + o_p(m^{-1/2}) \ , \\ U_x &= F_m^{o\ -1}((m - d_x + 1)/m) = \theta_x + Z_x/[2m^{1/2}\ell(0)] - \bar{\xi}_m(\theta_x)/\ell(0) + o_p(m^{-1/2}). \end{aligned} \tag{3.2}$$

Proof. By proposition 2,

$$\begin{aligned} L_x &= F^{o\ -1}(d_x/m) - \bar{\xi}_m(F^{o\ -1}(d_x/m))/\ell(F^{o\ -1}(d_x/m) - \theta_x) + o_p(m^{-1/2}), \\ &= \theta_x + [1/(2m) - Z_x/(2m^{1/2})]/\ell(0) - [\bar{\xi}_m(F^{o\ -1}(d_x/m)) - \bar{\xi}_m(\theta_x)]/\ell(0) \\ &\quad - \bar{\xi}_m(\theta_x)/\ell(0) + [1/\ell(0) - 1/\ell(F_o^{-1}(d_x/m) - \theta_x)]\ \bar{\xi}_m(F^{o\ -1}(d_x/m)) + o_p(m^{-1/2}) \ . \end{aligned}$$

Since $\bar{\xi}_m(t)$ has mean 0 for all t and the process $\{m^{1/2}\ \bar{\xi}_m(t),\ 0 \le t \le T\}$ converges weakly to a continuous path Gaussian process, it follows that

$$\bar{\xi}_m(F^{\circ\,-1}(d_x/m)) - \bar{\xi}_m(\theta_x) = o_p(m^{-1/2}) .$$

Also $\bar{\xi}_m(F^{\circ\,-1}(d_x/m)) = o_p(m^{-1/2})$ and $\ell(\cdot)$ is continuous at zero implies that the second to last term is also $o_p(m^{-1/2})$. We have thus proved L_x in (3.2) and L_y is similar. $\square$

Remark. The remainder $o_p(m^{-1/2})$ in (3.2) and (3.3) can be replaced by $o(m^{-1/2})$ a.s. with a modified proof or an even better rate (say $O((\ell n\ m/m)^{3/4})$ a.s.) if we further assume that the second derivative of L exists. While all the asymptotic results on convergence in probability in this section can be generalized to convergence a.s., we shall not attempt to do so as the current forms suffice for our purpose.

Statements similar to Theorem 1 apply to L_y and U_y as well with corresponding Z_y and d_y. Let

$$\tau_x = \int_0^{\theta_x} [\bar{F}(s)]^{-2}\, dF_1(s) \tag{3.3}$$

then $\mathrm{Var}(\bar{\xi}_m(\theta_x)) = \tau_x/(4m)$. The central limit theorem applied to $\bar{\xi}_m(\theta_x)$ then gives:

Corollary. As $m \to \infty$.

(i) $\quad m^{1/2}(L_x - \theta_x)$ converges weakly to a Normal distribution with mean $- Z_x/[2\ \ell(0)]$ and variance $\tau_x/[2\ \ell(0)]^2$.

(ii) $\quad m^{1/2}(U_x - L_x)$ converges in probability to $Z_x/\ell(0)$.

(iii) $P(\theta_x < L_x)$ <u>converges to</u> $\Phi(-Z_x/(\tau_x)^{1/2})$, <u>where Φ is the standard normal distribution function</u>.

The approximate confidence coefficient for $[L_x, U_x]$ is thus

$$\gamma_x = 1 - 2\Phi(-Z_x/(\tau_x)^{1/2}). \tag{3.4}$$

It should be noted that $\tau_x = 1$ in the uncensored case.

The statements in the Corollary also applies to L_y and U_y with $\tau_y = \int_0^{\theta_y} [\bar{G}(s)]^{-2} \, dG_1(s)$.

Let $N = n + m$, and assume that m/N tends to λ, $0<\lambda<1$, as m and n tend to infinity. The following theorem relates α, the overall error rate, to Z_x and Z_y which determine γ_x and γ_y.

<u>**Theorem 2**</u>. $P(U_y - L_x < \Delta) + P(L_y - U_x > \Delta) \to 2\Phi\left(-\dfrac{(1-\lambda)^{1/2} Z_x + \lambda^{1/2} Z_y}{[(1-\lambda)\tau_x + \lambda \tau_y]^{1/2}}\right)$.

<u>**Proof**</u>. By the Corollary, $N^{1/2}(U_y - \theta_y) - N^{1/2}(L_x - \theta_x)$ converges weakly to a normal distribution with mean $Z_y/[2\ell(0)(1-\lambda)^{1/2}] + Z_x/[2\ell(0)\lambda^{1/2}]$ and variance $\tau_y/[2\ell(0)(1-\lambda)^{1/2}]^2 + \tau_x/[2\ell(0)\lambda^{1/2}]^2$. Hence

$$P(U_y - L_x < \Delta)$$

$$= P(N^{1/2}(U_y - \theta_y) - N^{1/2}(L_x - \theta_x) < 0) \to \Phi\left(-\frac{\lambda^{1/2} Z_y + (1-\lambda)^{1/2} Z_x}{[\lambda \tau_y + (1-\lambda)\tau_x]^{1/2}}\right).$$

Similarly, $P(L_y - U_x > \Delta)$ converges to the same limit. $\square$

Define Z_α by $\alpha = 2\Phi(-Z_\alpha)$, and let Z_x, Z_y satisfy

$$(1-\lambda)^{1/2} Z_x + \lambda^{1/2} Z_y = Z_\alpha[(1-\lambda)\,\tau_x + \lambda\,\tau_y]^{1/2} \tag{3.5}$$

Then for depths d_x, d_y given by (3.1), we have

$$P(L_y - U_x < \Delta < U_y - L_x) \to 1-\alpha ,$$

and under H_0: $\Delta = 0$, $P(L_x > U_y) + P(L_y > U_x) \to \alpha$. Hence $[L_y - U_x, U_y - L_x]$ is a confidence interval for Δ with approximate confidence coefficient $(1-\alpha)$, and the test which rejects H_0 whenever 0 is not in the confidence interval, or equivalently whenever the two quantile-intervals are disjoint, has approximate size α.

The condition (3.5) provides an infinite number of choices for Z_x and Z_y. The following Theorem shows that the asymptotic length of any of the two-sample confidence intervals does not depend on Z_x and Z_y.

Theorem 3. Let Λ denote the length of the two-sample confidence interval. Then $N^{1/2}\Lambda \to Z_\alpha[(1-\lambda)\tau_x + \lambda\tau_y]^{1/2}/\{[\lambda(1-\lambda)]^{1/2}\ell(0)\}$ in probability.

Proof.

$$N^{1/2}\Lambda = N^{1/2}(U_y - L_y) + N^{1/2}(U_x - L_x)$$

$$\to Z_y/[(1-\lambda)^{1/2}\ell(0)] + Z_x/[\lambda^{1/2}\ell(0)] ,$$

by Corollary (ii). The theorem now follows from (3.5). $\square$

We now consider three ways to specify Z_x and Z_y.

1. Quantile-intervals with equal confidence coefficients:

If we choose $\gamma_x = \gamma_y$, then $Z_x/\tau_x^{1/2} = Z_y/\tau_y^{1/2}$. The condition (3.5) then implies

$$Z_x = \frac{\tau_x^{1/2}\,[(1-\lambda)\tau_x + \lambda\tau_y]^{1/2}}{[(1-\lambda)\tau_x]^{1/2} + [\lambda\tau_y]^{1/2}}\, Z_\alpha \approx \frac{\tau_x^{1/2}\,(n\tau_x + m\tau_y)^{1/2}}{(n\tau_x)^{1/2} + (m\tau_y)^{1/2}} \cdot Z_\alpha\,,$$

$$Z_y = (\tau_y/\tau_x)^{1/2}\, Z_x \approx \frac{\tau_y^{1/2}\,(n\tau_x + m\tau_y)^{1/2}}{(n\tau_x)^{1/2} + (m\tau_y)^{1/2}} \cdot Z_\alpha \qquad (3.6)$$

2. Quantile-intervals with equal asymptotic lengths:

In this case, $m^{-1/2}\, Z_x = n^{-1/2}\, Z_y$. Combining this with condition (3.5), we have

$$Z_x = \frac{[(1-\lambda)\tau_x + \lambda\tau_y]^{1/2}}{[m(1-\lambda)]^{1/2} + (n\lambda)^{1/2}} \cdot m^{1/2}\, Z_\alpha \approx \frac{(n\tau_x + m\tau_y)^{1/2}}{2n^{1/2}} \cdot Z_\alpha$$

$$Z_y = (n/m)^{1/2}\, Z_x \approx \frac{(n\tau_x + m\tau_y)^{1/2}}{2\, m^{1/2}} \cdot Z_\alpha \qquad (3.7)$$

3. Quantile intervals with equal depths:

In this case $d_x = d_y$, and (3.1) implies $m - n = m^{1/2}\, Z_x - n^{1/2}\, Z_y$. Condition (3.5) now gives

$$Z_x = \frac{[(1-\lambda)\tau_x + \lambda\tau_y]^{1/2}\, n^{1/2}\, Z_\alpha + \lambda^{1/2}(m-n)}{[n(1-\lambda)]^{1/2} + [m\lambda]^{1/2}}$$

$$\approx \frac{(n\tau_x + m\tau_y)^{1/2} n^{1/2} Z_\alpha + m^{1/2}(m-n)}{n+m}$$

$$Z_y = (m/n)^{1/2} Z_x - (m-n)n^{-1/2} \approx \frac{(n\tau_x + m\tau_y)^{1/2} m^{1/2} Z_\alpha - (m-n)n^{1/2}}{n+m} . \tag{3.8}$$

Remarks. 1. When the sample sizes are equal, (3.8) reduce to (3.7) and in both cases $Z_x = Z_y = Z_\alpha (\tau_x + \tau_y)^{1/2}/2$. However, unlike the uncensored case (where $\tau_x = \tau_y = 1$), (3.6) does not reduce to (3.7).

2. Under $\tau_x = \tau_y$, the Z_x in (3.6) and (3.7) equals $\tau_x^{1/2}$ times the Z_x for the corresponding uncensored case, but in (3.4) the $\tau_x^{1/2}$ cancels and the confidence coefficients γ_x and γ_y remain the same as the uncensored case. Neither Z_x, Z_y nor γ_x, γ_y produced by (3.8) remains the same as the uncensored case. If in addition $m = n$, all three solutions reduce to $Z_x = Z_y = Z_\alpha \tau_x^{1/2}/2$. Note that if the censoring schemes are different for the two samples, in general $\tau_x \neq \tau_y$ even under H_0: $\Delta = 0$.

3. It should be noted here, as implied by Theorem 1 and (3.5), that the implementation of the confidence interval and test for Δ rely only on τ_x and τ_y and not on the density at the median. Hence our procedures avoid estimating the density at the median which is required by Brookmeyer and Crowley (1982ab) and Wei and Gail (1983). All procedures essentially require the consistent estimation of τ_x and τ_y. In practice this can be done in several ways, for example,

$$\hat{\tau}_x = \sum_{X_i \le m_x} \frac{md_i}{N(X_i)[N(X_i) + d_i]}, \tag{3.9}$$

where $m_x = F_m^{o\,-1}(1/2)$ is the x-sample median, d_i is the number of observed deaths at X_i and $N(X_i) = \sum_{j=1}^{m} I(X_j > X_i)$ is the number of observed survival times larger than X_i. The estimator $\hat{\tau}_x$ is similar to formula (2) of Brookmeyer & Crowley (1982a). See also (5.2) of Brookmeyer and Crowley (1982b) for an alternative estimate.

4. Mood's median test was shown to be a special case produced by sign-intervals in Hettmansperger (1984a). With censoring, the median test statistics can no longer be expressed in terms of the differences of order statistics as in formula (6) there. It is not clear whether the median test by either Prentice (1978) or Brookmeyer and Crowley (1982b) can be produced by our quantile-intervals.

Section 4. Tables for confidence coefficients and a recommendation

Confidence coefficients γ_x and γ_y corresponding to a two-sample test of H_0: $\Delta = 0$ versus H_A: $\Delta \ne 0$ at level $\alpha = .05$ will be provided in this section for the three intervals described in the previous section. Due to censoring, the confidence coefficients γ_x and γ_y depend on the value of τ_x and τ_y so we shall only consider a special case, the proportional hazards model, where τ_x and τ_y can be evaluated easily. Let $\bar{H}(t) = [\bar{F}^o(t)]^r$, $\bar{K}(t) = [\bar{G}^o(t)]^s$, then $\tau_x = (2^{r+1} - 1)/(r + 1)$ and π_x = probability of x-censoring = $r/(r + 1)$. Similarly, $\tau_y = (2^{s+1} - 1)/(s + 1)$ and π_y = probability of y-censoring = $s/(s + 1)$. We shall consider r, s = 1/3, 1, 3, which correspond to 25%, 50% and 75% censoring.

(Tables 1-3 here)

Note that $\tau_x = \tau_y$ when $r = s$, then (3.6) and (3.7) give the same γ_x and γ_y as in the uncensored case. Hence, the first two columns which corresponds to $r = s$ in Tables 1 and 2 should be the same as those columns corresponding to equal-coefficient and equal - length case in Table 1 of Hettmansperger (1984a). The discrepancy is due to computational accuracy. In the equal depth case (Table 3), (3.8) depends not only on the ratio of sample sizes but also the values of m. We chose m = 90 so that the corresponding n is an integer. Our computation shows that the pair of coefficients are very high (over .95 for $m/n \geq 1.5$ and close to 1 for $m/n \geq 2$) for unequal sample sizes. Therefore we only provide the results for ratio of sample sizes (m/n) that are 1 and 1.5 in Table 3.

As can be seen from the tables, the confidence coefficients in Table 1 appear to be quite stable (between .83 and .88) for various ratio of sample sizes and degrees of censoring compared to those in Tables 2 and 3. For this reason the equal coefficients solution (3.6) seems preferable to the others and we recommend its use in applications. In an unpublished paper of Tableman and Hettmansperger (1988) the equal lengths case may be preferable in some uncensored cases based on Bahadur efficiency. In the censored case, the equal lengths solution (3.7) may be preferable to the equal coefficients solution (3.6) since one only needs to estimate $\tau_x + \tau_y$ when m = n or $(1-\lambda)\tau_x + \lambda\tau_y$ in general, while in (3.6) one has to estimate, in addition, τ_x and τ_y separately. The choices of Z_x and Z_y (or equivalently γ_x and γ_y) for (3.8) are not recommended since they are more complicated than (3.7) even when $\tau_x = \tau_y$ and they require much higher coefficients (see Table 3) for unbalanced sample sizes compared to those coefficients provided in Tables 1 and 2.

Section 5. An Example

We shall apply the two-sample procedure in section 3 to data from a Phase III colorectal cancer clinical trial (Ansfield and Klotz 1977). The data analyzed were an updated version reported in the Central Oncology Group Final Report (COG 7030), Spring 1977. These data were also analyzed by Brookmeyer and Crowley (1982a) to compare confidence intervals for the median survival times of four dosage regimens 5-fluorouracil. Refer to Brookmeyer and Crowley (1982a) for a description of the data. A summarization of the data is given in Table 4 together with 95% confidence intervals. The median survival time (weeks) and the quantile-intervals are computed by a continuous version of the Kaplan-Meier estimator, denoted by F_n^*, discussed in Remark 1 of section 6 and is therefore slightly different from the median in Table 7 of their paper.

(Table 4 here)

For $\gamma_x = .95$ the Z_x in (3.4) is equal to 1.96 times $(\tau_x)^{1/2}$ where τ_x is estimated by $\hat{\tau}_x$ in (3.9). The depth d_x in Table 4 is then obtained from (3.1) which for treatment 1 is 19.09. The quantile-interval $[L_x, U_x)$ is obtained by $L_x = F_m^{*-1}(d_x/m)$ and $U_x = F_m^{*-1}((m - d_x + 1)/m)$, where F_m^* is the linear interpolated Kaplan-Meier estimate as discussed in Remark 1 of section 6. A table for the survival estimate $\hat{S}^o = 1 - F_m^o$ is available for treatment 1 in Table 8 of Brookmeyer and Crowley (1982a). Our estimate F_m^o (which is not included here) coincides with theirs.

To make pairwise comparisons at the 5 per cent significance level we need the confidence coefficients γ_x and γ_y and the depth values d_x and d_y. Table 5 gives the results for the three different methods to specify Z_x and Z_y (cf. formula (3.6) to (3.8)).

(Table 5 here)

It turned out that none of the one-sample quantile intervals in Table 5 are disjoint hence all three methods fail to detect a significant difference within any of the six pairs. Brookmeyer and Crowley (1982b, section 6) applied the k-sample (k=4) median test to this data set and also found no significant differences among the four treatments. Our conclusion is therefore in agreement with theirs. It is worth noting here that the quantile-intervals with equal asymptotic lengths actually yield intervals with quite different lengths. This suggests that higher order accuracy is needed for the asymptotic length than that given in Corollary (ii) of section 3. In our current example the sample sizes (and τ_x) for each treatment are fairly even, hence the quantile-intervals produced by all three methods ((3.6) to (3.8)) are comparable.

Section 6. Discussion and generalization

1. The quantile-intervals that are defined in (2.1) and (2.2) are based on the PLEs F_m^o and G_n^o which are step functions. Since the distribution functions F^o and G^o are assumed to be continuous it may be preferable to use a continuous version (for example, linear interpolation) of the PLE. Let F_m^* and G_m^* denote such a version by connecting the jump points (c.f. Figure 1 of Brookmeyer and Crowley (1982b)). It follows from the appendix of Gastwirth and Wang (1988) that

$$\sup_{o \leq t \leq T} |F_m^*(t) - F_m^o(t)| = O(m^{-1}) \quad \text{a.s.} \tag{6.1}$$

and

$$\sup_{o<p\le p_o} |F_m^{o\,-1}(p) - F_m^{*\,-1}(p)| = o_p(m^{-1/2}) \quad \text{a.s.}, \tag{6.2}$$

for any $o<p_o<1$. In practice, one may use the smoothed quantile-interval $[L_x^*, U_x^*]$ by replacing F_m^o by F_m^* in (2.1). It is immediate from (6.2) that all the asymptotic results in section 3 apply to the smoothed quantile-intervals $[L_x^*, U_x^*]$.

2. Theorem 3 shows that all the intervals have the same asymptotic length and $N^{1/2}\,\Lambda\,/(2\,Z_\alpha)$ converges in probability to $[(1-\lambda)\tau_x + \lambda\tau_y]^{1/2}/[4\lambda(1-\lambda)\ell^2(0)]^{1/2}$ which is the reciprocal of the Pitman efficacy (c.f. Hettmansperger (1984b)) of the control median test by Gastwirth and Wang (1988). It can be checked easily that both median tests by Brookmeyer and Crowley (1982b) and Prentice (1987) have the same efficacy as the control median test. Hence all four median type tests have the same efficiency.

3. The two-sample inference method based on quantile-intervals was developed only under the location model. For the Behrens-Fisher problem where we test whether both populations have the same median but may differ in shape, the significance level

$$\alpha \to \Phi\left(- \frac{f^o(\theta_x)Z_y\,\lambda^{1/2} + g^o(\theta_y)Z_x(1-\lambda)^{1/2}}{[\lambda\tau_y[f^o(\theta_x)]^2 + (1-\lambda)\tau_x[g^o(\theta_y)]^2\}^{1/2}}\right) .$$

Then one must estimate, in addition to τ_x and τ_y, the ratio $f^o(\theta_x)/g^o(\theta_y)$, where f^o and g^o are the densities of F^o and G^o, respectively. The estimation of $f^o(\theta_x)/g^o(\theta_y)$ can be accomplished by density estimation procedures in the censored case (Lo, Mack and Wang (1986), Padgett and McNichols (1984)) which usually require very large sample sizes. Hence, the procedure in this paper can be used for the Behrens-Fisher problem as in Hettmansperger (1973).

4. We demonstrated in this paper how to conduct two-sample inference methods based on one-sample procedures. Although this is done for the quantile-intervals (or sign-tests) under the location model it can be applied to other nonparametric procedures (e.g. linear rank tests) and models (e.g. scale model or stochastically ordered alternative). The crucial step lies in the asymptotic distribution of the endpoints of the one-sample confidence intervals (cf. Corollary in section 2).

Acknowledgement: The authors wish to thank Paul Johnson for his most helpful computational assistance.

Table 1. Equal Confidence Coefficients to Produce a Size .05 Two-sided Two-Sample Test ($\gamma_x = \gamma_y$)

m/n	$\pi_x = \pi_y$	(.25,.5)	(.25,.75)	(.5,.25)	(.5,.75)	(.75,.25)	(.75,.5)
1	.834	.835	.851	.835	.844	.851	.844
1.5	.836	.84	.862	.834	.854	.842	.837
2	.840	.846	.870	.836	.862	.837	.835
2.5	.844	.851	.876	.839	.868	.835	.843
3	.849	.855	.881	.843	.873	.834	.835

(Column groups other than m/n are values of (π_x, π_y).)

Table 2. Confidence Coefficients Corresponding to Equal Lengths

m/n	$\pi_x = \pi_y$		(.25,.5)		(.25,.75)		(.5,.25)		(.5,.75)		(.75,.25)		(.75,.5)	
	γ_x	γ_y	γ_x	γ_y	γ_x	γ_y	γ_x	γ_y	γ_x	γ_y	γ_x	γ_y	γ_x	γ_y
1	.834	.834	.864	.806	.958	.737	.806	.864	.933	.754	.737	.958	.754	.933
1.5	.879	.794	.909	.771	.983	.717	.848	.821	.967	.930	.763	.920	.785	.890
2	.910	.770	.938	.750	.993	.707	.880	.793	.984	.717	.786	.889	.811	.858
2.5	.933	.754	.958	.737	.997	.701	.905	.774	.992	.709	.806	.864	.834	.834
3	.950	.742	.971	.727	.999	.696	.924	.76	.996	.703	.826	.844	.854	.815

Table 3. Confidence Coefficients with Equal Depth ($d_x = d_y$)

(π_x, π_y)	m=n=90		m=90, n=60	
	γ_x	γ_y	γ_x	γ_y
(.25,.25)	.834	.834	.997	.997
(.25,.5)	.864	.806	.998	.993
(.25,.75)	.958	.737	.999	.963
(.5 ,.25)	.806	.864	.993	.998
(.5 ,.5)	.834	.834	.995	.995
(.5 ,.75)	.933	.754	.999	.966
(.75,.25)	.737	.958	.946	.999
(.75,.5)	.754	.933	.951	.998
(.75,.75)	.834	.834	.973	.979

Table 4. Summary Statistics and 95% Confidence Intervals for Four Treatments

Treatment	1	2	3	4
Sample size	53	56	58	52
Censored proportion	16/53	8/56	14/58	7/52
Median survival time (weeks)	59.18	39.53	43.23	28.52
Estimated τ_x	1.109	1.152	1.23	.90
Depth d_x	19.09	20.05	20.33	20.15
95% quantile-interval $[L_x, U_x)$	[37.4,72.5)	[28.6,52.3)	[25.9,61.4)	[24.3,45.0)
95% confidence interval by Brookmeyer and Crowley	[38,73)	[31,51)	[28,60)	[25,46)

Table 5. Two-Sample Procedures for all Six Pairwise Comparisons, $\alpha = 0.05$

	Equal coefficients		Equal lengths		Equal depths	
Treatment 1 & 2						
γ_x γ_y	.834	.834	.832	.836	.780	.878
d_x d_y	21.69	22.93	21.71	22.91	22.29	22.29
$[L_x, U_x)$	[39.6, 64.7)		[39.6, 64.6)		[40.3, 63.3)	
$[L_y, U_y)$	[34.8, 48.2)		[34.8, 48.3)		[34.2, 49.4)	
Treatment 1 & 3						
γ_x γ_y	.834	.834	.835	.833	.746	.897
d_x d_y	21.69	23.65	21.67	23.67	22.62	22.62
$[L_x, U_x)$	[39.6, 64.7)		[39.6, 64.8)		[41.1, 62.6)	
$[L_y, U_y)$	[35.5, 51.6)		[35.5, 51.6)		[32.0, 58.2)	
Treatment 1 & 4						
γ_x γ_y	.835	.835	.814	.855	.832	.838
d_x d_y	21.68	21.76	21.92	21.52	21.72	21.72
$[L_x, U_x)$	[39.6, 64.7)		[39.8, 64.0)		[39.6, 64.6)	
$[L_y, U_y)$	[24.7, 42.0)		[24.6, 42.4)		[24.7, 42.1)	
Treatment 2 & 3						
γ_x γ_y	.834	.834	.831	.837	.859	.806
d_x d_y	23.65	22.93	23.69	22.89	23.29	23.29
$[L_x, U_x)$	[34.8, 48.2)		[34.8, 48.3)		[35.1, 47.6)	
$[L_y, U_y)$	[35.5, 51.6)		[35.5, 51.5)		[35.1, 52.2)	
Treatment 2 & 4						
γ_x γ_y	.835	.835	.816	.853	.876	.778
d_x d_y	21.76	22.93	21.54	23.16	22.32	22.32
$[L_x, U_x)$	[34.8, 48.3)		[35.0, 47.8)		[34.2, 49.3)	
$[L_y, U_y)$	[24.7, 42.0)		[24.6, 42.4)		[24.8, 40.3)	
Treatment 3 & 4						
γ_x γ_y	.835	.835	.813	.856	.895	.740
d_x d_y	23.64	21.76	23.93	21.50	22.65	22.65
$[L_x, U_x)$	[35.5, 51.6)		[35.7, 51.1)		[32.2, 58.2)	
$[L_y, U_y)$	[24.7, 42.0)		[24.6, 42.5)		[24.9, 38.9)	

References

ANSFIELD, F., KLOTZ, J., and THE CENTRAL ONCOLOGY GROUP (1977), "A Phase III Study Comparing the Clincial Utility of Four-Dosage Regimens of 5-Fluorouracil," Cancer, 39, 34-40.

BROOKMEYER, R. and CROWLEY, J. (1982a), "A Confidence Interval for the Median Survival Time," Biometrics, 38, 29-41.

BROOKMEYER, R. and CROWLEY, J. (1982b), "A K-sample Median Test for Censored Data," Journal of the American Statistical Association, 77, 433-440.

EFRON, B. (1967), "The Two-sample Problem with Censored Data," Proceeding of the 5th Berkeley Symposium, 4, 831-853.

GASTWIRTH, J.L. and WANG, J.L. (1988), "Control Percentile Test Procedures for Censored Data," Journal of Statistical Planning and Inference, 18, in press.

GEHAN, E. (1965), "A Generalized Wilcoxon Test for Comparing Arbitrary Singly-Censored Samples," Biometrika, 52, 203-223.

HETTMANSPERGER, T.P. (1973), "A Large Sample Conservative Test for Location with Unknown Scale Parameter," Journal of the American Statistical Association, 68, 466-468.

HETTMANSPERGER, T.P. (1984a), "Two-sample Inference Based on One-sample Sign Statistics," Applied Statistics, 33, No. 1, 45-51.

HETTMANSPERGER, T.P. (1984b), Statistical Inference Based on Ranks, New York: Wiley.

HODGES, J.L., Jr. and LEHMANN, E.L. (1963), "Estimates of Location Based on Ranks," Annals of Mathematical Statistics, 34, 598-611.

KAPLAN, E.L. and MEIER, P. (1958), "Nonparametric Estimation from Incomplete Observations," Journal of the American Statistical Association, 53, 457-481.

LO, S.H., MACK, Y.P., and WANG, J.L. (1986), "Density and Hazard Rate Estimation for Censored Data via Strong Representation of the Kaplan-Meier Estimator," Technical Report, University of California, Davis. Tentatively accepted by Probability Theory and Related Fields.

LO, S.H. and SINGH, K. (1986), "The Product Limit Estimator and the Bootstrap: Some Asymptotic Representations", Probability Theory and Related Fields, 71, 455-465.

MOOD, A.M. (1950), Introduction to the Theory of Statistics, New York: McGraw-Hill.

PADGETT, W.J. and McNICHOLS, D.T. (1984), "Nonparametric Density Estimation from Censored Data," Communications in Statistics, Ser. A, 13, 1581-1611.

PRENTICE, R.L. (1978), "Linear Rank Tests with Right Censored Data," Biometrika, 65, 167-179.

REID, N. (1981), "Estimating the Median Survival Time," Biometrika, 68, 601-608.

SERFLING, R.J. (1980), Approximation Theorems of Mathematical Statistics, New York: Wiley.

TABLEMAN, M. and HETTMANSPERGER, T.P. (1988), "Bahadur Efficiency of Two-Sample Tests Based on One-Sample Sign Statistics," Journal of Royal Statistical Society (B), to appear.

WEI, L.J. and GAIL, H.H. (1983), "Nonparametric Estimation for a Scale-change with Censored Observations," Journal of the American Statistical Association, 78, 382-388.

Unclassified

REPORT DOCUMENTATION PAGE — READ INSTRUCTIONS BEFORE COMPLETING THIS FORM

1. REPORT NUMBER: 74	2. GOVT ACCESSION NO. / 3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle): Two-Sample Inference for Median Survival Times Based on One-Sample Procedures for Censored Survival Data	5. TYPE OF REPORT & PERIOD COVERED
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): Jane-Ling Wang, The Wharton School, University of Pennsylvania; Thomas P. Hettmansperger, The Pennsylvania State University	8. CONTRACT OR GRANT NUMBER(s): N00014-80-C-0741
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Department of Statistics, The Pennsylvania State University, University Park, PA 16802	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: NR042-446
11. CONTROLLING OFFICE NAME AND ADDRESS: Office of Naval Research, Statistical and Probability Program Code 436, Arlington, VA 22217	12. REPORT DATE: August 1988
	13. NUMBER OF PAGES: 23
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report): Unclassified
	15a. DECLASSIFICATION DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Product-limit estimator, median test, sign test, confidence interval, quantile, asymptotic relative efficiency.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Confidence intervals for median survival times are derived for censored survival data. The intervals are obtained by using the quantiles of the Kaplan-Meier product limit estimator and have the same Pitman efficiency as the intervals found by inverting the sign test. Two-sample tests and confidence intervals for the difference in median survival times are then developed based on the comparison of the one-sample confidence intervals. Several methods for choosing the confidence coefficients of the corresponding one-sample confidence intervals are developed. The Pitman efficiencies

DD FORM 1473 EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Block 20 (Abstract Continued):

of these two-sample tests are the same as that of the median test proposed by Brookmeyer and Crowley (1982b). The procedures can also be used for the Behrens-Fisher problem.

Nonparametric two-sample inference procedures are useful in comparing the responses of treatment and control groups. In medical follow up studies the data are usually subjected to censoring. While many of the two-sample procedures have been extended to accommodate censored data we show, in this paper, how to construct two-sample tests and confidence intervals based on one-sample confidence intervals. This method was first discussed in Hettmansperger (1984a) for uncensored data where a confidence interval for the difference in population medians is constructed by subtracting the endpoints of one (one-sample) interval from the opposite endpoints of the other. The test then rejects the null hypothesis of equal medians if the one-sample intervals are disjoint. Hettmansperger used the sign-interval as the one-sample interval which is obtained by inverting the sign-test. In the presence of censoring we modify the sign-interval to the so called quantile-interval whose endpoints are the quantiles of the product-limit estimator of Kaplan and Meier (1958). After deriving the asymptotic properties of the quantile-interval we show, for specified overall level α, three ways to select the confidence coefficients for the one-sample quantile intervals. For $\alpha = 0.05$ and equal confidence coefficients for both samples, the one-sample confidence coefficients are in the neighborhood of .85 under the proportional hazard model. The procedures are then applied to data from a colorectal cancer clinical trial to compare four treatments.